DOCUMENT RESUME

ED 278 716 TM 870 146

AUTHOR Marsh, Herbert W.
TITLE Multidimensional Self Concepts: Do Positively and Negatively Worded Items Measure Substantively Different Components of Self.
PUB DATE [86]
NOTE 16p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Behavior Rating Scales; *Construct Validity; Correlation; Foreign Countries; Higher Education; High Schools; Multiple Regression Analysis; Personality Assessment; *Self Concept; *Self Concept Measures; Significant Others; *Test Construction; *Test Items; *Test Theory; Weighted Scores
IDENTIFIERS Australia; *Self Description Questionnaire

ABSTRACT

The purpose of the present investigation was to develop a construct validity approach for testing whether the separation of positive and negative item subscales is substantively meaningful in self-concept research. Results from three published studies using the Self Description Questionnaire (SDQ) III were reanalyzed. The SDQ III measures 13 distinct areas of self-concept and half of the items are negatively worded. In the reanalyses, no support was found for differentially weighting responses to positive and negative item subscales for any of the scales. Differential weighting produced little or no improvement in the prediction of self-concepts inferred by significant others (Study 1), in the short-term stability of responses (Study 2), in the consistency of responses across split-halves of each set of items (Study 3), or in the prediction of mathematical and verbal achievement (Study 3). The lack of support for the separation of positive and negative item subscales for reliability-like criteria such as stability and internal consistency suggests that support is unlikely for any validity criteria. Because scales constructed with negatively and positively worded items are a special case of the more general bipolar rating scale, the methodological approach has broad applicability for personality research. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED278716

Multidimensional Self Concepts: Do Positively and Negatively Worded Items Measure Substatively Different Components of Self

Herbert W. Marsh
University of Sydney, Australia

Running Head: Multidimensional Self Concepts

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. W. Marsh

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM 870 146

BEST COPY AVAILABLE

Multidimensional Self Concepts: Do Positively and Negatively Worded Items Measure Substatively Different Components of Self

Abstract

The purpose of the present investigation was to more fully develop a construct validity approach for testing whether the separation of positive and negative item subscales is substantively meaningful in self-concept research. In reanalyses of three studies, no support was found for differentially weighting responses to positive and negative item subscales for any of 13 self-concept scales. Differential weighting produced little or no improvement in the prediction of self-concepts inferred by significant others (Study 1), in the short-term stability of responses (Study 2), in the consistency of responses across split-halves of each set of items (Study 3), or in the prediction of mathematical and verbal achievement (Study 3). The lack of support for the separation of positive and negative item subscales for reliability-like criteria such as stability and internal consistency suggests that support is unlikely for any validity criteria. Because scales constructed with negatively and positively worded items are a special case of the more general bipolar rating scale, the methodological approach has broad applicability for personality research.

Multidimensional Self Concepts: Do Positively and Negatively Worded Items Measure Substatively Different Components of Self

Test construction specialists argue that the number of positively and negatively worded items should be balanced on personality, attitude and other rating scales in order to disrupt possible response sets. This recommendation assumes that positively and negatively worded items measure the same construct, but this assumption is rarely tested and its validity may be dubious in some situations. Benson and Hocevar (in press) and Marsh (1986), for example, argued that negatively worded items produce systematic biases in the responses by young children. Marsh showed that this influence on responses to a multidimensional self-concept instrument was age-related, and related to verbal skills when age was held constant. He concluded that the influence of negative items was a response bias that had a cognitive developmental basis, and recommended against the use of negatively worded items for young children.

More generally, the tendency for subjects to respond to personality rating items independently of item content is referred to as response set, response bias, response style, or a method/halo effect. Different approaches emphasize the nonsubstantive or substantively irrelevant components of responses to structured items (see Wiggins, 1973 for a review). Jackson (1967; Jackson & Messick, 1958, 1961) argued that content is what is left over after sources of style and method have been removed through approaches such as regression and factor analysis. A negative-item bias produced by cognitive developmental influences clearly qualifies as a response bias, but it is not the only basis for the effect of negatively worded items. Older respondents are generally able to cope with the cognitive demands of negatively worded items, but researchers still report that positively and negatively worded items designed to measure the same construct are empirically distinct (e.g., Bachman & O'Malley, in press; Carmines & Zeller, 1978; Gaudry & Spielberger, 1971; Marsh & Myers, 1986; Naylor, 1978; Russell & Antill, 1984). However, because the basis of the distinction between such positive and negative item factors is not easily identified, there is ambiguity in deciding whether the distinction is substantively important or a substantively irrelevant artifact of response bias. This question is the focus of the present investigation.

The Construct Validity Approach

Carmines and Zeller (1979; also see Bachman & O'Malley, in press) found that positively and negatively worded items from the Rosenberg (1965) self esteem scale define clearly distinguishable factors and he proposed two

alternative interpretations: (a) there are separate components of positive and negative self-esteem that are substantively meaningful; and (b) the separation of the two factors is an artifact of nonrandom measurement error. Using a construct validity approach, they reasoned that if the positive and negative item subscales measure substantively different constructs then they should be differentially related to external criteria, whereas if the distinction is due to an artifact then they should be similarly related to external criteria. Because positive and negative item subscales were similarly related to a set of external criteria, Carmines and Zeller concluded that the empirically distinguishable factors were "a function of a single theoretical dimension of self-esteem that is contaminated by a method artifact, response set" (p. 69).

The purpose of the present investigation is to develop more fully the construct validity approach for distinguishing between the alternative interpretations posited by Carmines and Zeller. There are at least two sets of problems with their approach to this issue. First, they based their conclusions on the observation that positive and negative item subscales were similarly correlated to external criteria, but this comparison may be uninterpretable: (a) the similarity in the correlations does not mean that each of the factors accounts for the same criterion variance, nor that some empirically weighted combination of the two would not perform substantially better than either considered separately or their unweighted combination; and (b) systematic differences in the correlations do not mean that the separation of positive and negative item subscales is substantively justified, because this would be expected if, for example, the negative wording produced a systematic method effect that was independent of external criteria. A better approach, one that is consistent with Carmines and Zeller's logic if not their analyses, is to compare empirically weighted and unweighted combinations of the positive and negative item subscales. If the unweighted sum of the positive and negative scales is as highly related to appropriate validity criteria as an empirically optimum weighting of the two subscales, then it is unlikely that the separation of the two subscales is substantively important.

The second set of potential problems concerns the generalizability Carmines and Zellers' findings to self-concept research. They correlated their positive and negative item subscales with "a set of theoretically relevant external variables" (p. 67), but in fact most of these variables were nearly uncorrelated with either esteem subscale (only 1 of 16 variables correlated higher than .25 with either subscale). A better approach would be to use more carefully selected validity criteria that were more

systematically related to esteem on the basis of theory and previous research. Furthermore, Carmines and Zeller's conclusions were based on a single dimension of self-concept as measured by one specific set of items. More recently self-concept researchers have emphasized the multidimensionality of self-concept (e.g., Byrne, 1984; Marsh, Barnes & Hocevar, 1985; Marsh & Shavelson, 1985). Even though Carmines and Zeller emphasized their methodological approach rather than the generality of their findings to other self-concept research, there is a need to test the generality of their results with different criteria and a different self-concept instrument.

The Present Investigation

The purpose of the present investigation is to develop more fully the construct validity approach and to further test Carmines and Zeller's conclusions that positive and negative item subscales in self-concept research do not differ substantively. In order to accomplish this the results from three published studies using the Self Description Questionnaire (SDQ) III were reanalyzed. The SDQ III measures 13 distinct areas of self-concept and half of the 10 or 12 items used to infer each facet are negatively worded. In the first study (Marsh, Barnes, & Hocevar, 1985) respondents completed the SDQ III and significant others inferred the self-concepts of the respondents with the same instrument. For this study, the generality (validity) of the influence of positively and negatively worded items was tested across self-responses and responses-by-others. In the second study (Marsh, Richards & Barnes, 1986), respondents completed the SDQ III at the beginning and end of a one month control interval. For this study, the generality (stability) of the influence of positively and negatively worded items was tested over time. In the third study (Marsh & O'Neill, 1984) respondents completed the SDQ III and standardized measures of verbal and mathematical achievement were obtained. For this study, the positive and negative items for each scale were divided in half and the generality (internal consistency) of the influence of positively and negatively worded items was tested over these split halves. In the third study, the relation between academic achievement and academic self-concepts based on positively and negatively worded items was also examined.

Methods

The Self Description Questionnaire (SDQ) III.

The present investigation is based on responses to the SDQ III. The theoretical rationale for its construction, the actual items, and a considerable body of empirical support for its dimensionality, reliability, and validity are summarized elsewhere (Marsh, 1985; Marsh, Barnes & Hocevar,

1984; Marsh & Jackson, in press; Marsh & O'Niell, 1984; Marsh, Richards & Barnes, 1986; Marsh & Shavelson, 1985). Each of the 13 SDQ III scales is represented by 10 or 12 items, half of which are negatively worded, and subjects respond on an eight-point response scale that varies from "1-Definitely False" to "8-Definitely true." The 13 scales are: Physical Ability; Physical Appearance; Opposite Sex Relations; Same Sex Relations; Relations With Parents; Spiritual Values/Religion; Honesty; Emotional Stability; Verbal; Math; General Academic; Problem Solving; and General-Self (the General-Self scale was based on the Rosenberg scale that was the basis of Carmines and Zeller's research).

The Samples.

Data came from three previously published studies:

1) Study 1 consisted of the sample of 151 Australian university students (mean age = 21.9, 79% female) described by Marsh, Barnes and Hocevar (1985; also see Marsh & O'Niell, 1984, Study 2). As part of that study, subjects asked the person who knew them the best to complete the SDQ III as if they were the person who had given it to them (i.e., they were to predict what the subject had said).

2) Study 2 consisted of the 361 Outward Bound participants (mean age = 21.3, 76% males) described by Marsh, Richards and Barnes (1986; in press). As part of that study, participants completed the SDQ III one month before, the first day of, the last day of, and 18 months after the completion of a 26-day residential program. For purposes of the present investigation, only the Time 1 and 2 responses that were collected before the program intervention were considered.

3) Group 3 consisted of the 296 year-11 girls (mean age = 16.7) from two private catholic girls schools described by Marsh and O'Niell (1984). In addition to the self-concept responses, standardized measures of academic achievement were considered in that study.

Statistical Analysis

For all three studies there were: two sets of 13 scale scores representing an unweighted average of responses to positively and negatively worded items (T1 and T2); two sets of 13 subscale scores representing just positively worded items (P1 and P2); and two sets of 13 subscale scores representing just negatively worded items (N1 and N2). The two sets of scores consisted of: self-responses and responses-by-others (Study 1); responses by the respondents at Times 1 and 2 (Study 2); and responses by the respondents to items in the first and second halves of the

SDQ III (Study 3). For all three studies: simple correlation was used to relate T1 to T2, P1 to P2, and N1 to N2; multiple regression was used to relate T1 to the optimally weighted combination of P2 and N2, and T2 to the optimally weighted combination of P1 and N1; and canonical correlation was used to relate the optimally weighted combination of P1 and N1 to the optimally weighted combination of P2 and N2. The canonical correlation must necessarily be as large as or larger than either of the multiple Rs, and the multiple Rs must necessarily be as large or larger than the correlation between T1 and T2. To the extent that these differences are trivial, differentially weighting the positive and negative item subscales makes no difference and the results would provide no substantive support for the separation of the subscales. To the extent that these differences are large, there would be support for the separation of the subscales.

Results and Discussion.

Study 1.

In Study 1 each respondent completed the SDQ III and a significant other inferred the multiple self-concepts of the same person on SDQ III. Previous research (Marsh, Barnes & Hocevar, 1985) demonstrated that there was good self-other agreement based on total scores for each of the 13 self-concept scales[1]. The purpose of this reanalysis was to determine if this agreement was better when positive and negative item subscales of each scale were considered separately or differentially weighted.

Agreement on the 13 total scores for self-responses and responses-by-others (.57; see T1 with T2 in Table 1 for Study 1) is modestly higher than agreement on the 13 positive item subscales (.54) and the 13 negative item subscales (.52). More importantly, this agreement on the 13 total scores (.57) was virtually unimproved by differentially weighting the positive and negative item subscales for the self-responses, for the responses-by-others, or for both. The average agreement was .57, .58 and .59 respectively when the positive and negative item subscales were differentially weighted for the responses-by-others (T1 with P2 and N2), for self-responses (T2 with P1 and N1), and for both self-responses and responses-by-others (P1 and N1 with P2 and N2). It is important to reiterate that agreement on the (unweighted) total scores represents an absolute lower bound for agreement on the differentially weighted components[2]. Hence, differentially weighting the subscales produced a surprisingly small improvement, and this lack of improvement was consistent across all 13 SDQ III scales. These results provide no support for the separation of any of the 13 self-concept scales into positive and negative item subscales.

BEST COPY AVAILABLE

Insert Table 1 About Here

Study 2.

In Study 1 no support was found for the separation of positive and negative item subscales. However, the conclusions were based on agreement between self-responses and responses-by-others, and support for the separation of the subscales might be found if other validity criteria were used. Because it is impossible to test the generality of the conclusions with all potential validity criteria, an alternative approach was sought. It may be possible to find a criterion for which the lack of support would be so convincing that it would be unlikely that support would be found for other criteria. Such a criterion, the stability of responses by the same person to the same instrument over a short period of time, is the basis of Study 2. It must be emphasized that support for the separation of positive and negative item subscales based on their short-term stability would not provide convincing support for the validity of their separation, but that the lack of support for their separation using this criterion would provide convincing evidence against the validity of their separation.

In Study 2 each respondent completed the SDQ III before and after a one-month control interval, and previous research (Marsh, Richards & Barnes, 1985; in press) demonstrated that there was good Time 1/Time 2 agreement on total scores for each of the 13 SDQ III scales. The purpose of this reanalysis was to determine if this agreement was improved when positive and negative item subscales were considered separately or differentially weighted.

Agreement on the 13 total scores for Time 1 and Time 2 (.87; see T1 with T2 in Table 1 for Study 2) is modestly higher than agreement on the 13 positive item subscales (.84) and 13 negative item subscales (.83). More importantly, this agreement on the 13 total scores (.87) was virtually unimproved by differentially weighting the positive and negative item components of either Time 1 or Time 2 responses. The average agreement was .87, .87, and .87 respectively when the positive and negative item subscales were differentially weighted for Time 2 responses, for Time 1 responses, and for both Time 1 and 2 responses. Hence, differentially weighting the scales produced no improvement, and this lack of improvement was consistent across all 13 SDQ III scales. These results provide no support whatsoever for the separation of any of the SDQ III scales into positive and negative item subscales.

Study 3.

The logic of Study 3 is similar to that of Study 2. In Study 2 the

short-term stability of the responses to the same items was considered in order to test the validity of the separation of the positive and negative items subscales. In Study 3 the consistency of responses across different sets of items administered at the same time was considered. As in Study 2, support for the separation of positive and negative item subscales based on their split-half consistency would not provide convincing support for the validity of their separation, but the lack of support of their separation would provide convincing against the validity of their separation.

In Study 3 each respondent completed the SDQ III only once, and the 10 or 12 items from each scale were divided into four subscales -- positive items from the first half of the SDQ III, negative items from the first half, positive items from the second half, and negative items from the second half. Thus, although the analyses are the same as those in Studies 1 and 2, the total scores (T1 and T2) and positive and negative item scores (P1, P2, N1 and N2) are each based on only half as many items as in Studies 1 and 2.

Agreement on the 13 total scores for first-half and second-half responses (.77; see T1 with T1 in Table 1 for Study 3) was modestly higher than agreement on the 13 sets of positively worded items (.70) and negatively worded items (.62). More importantly, this agreement on the 13 total scores (.77) was virtually unchanged by differentially weighting the positive and negative item components of either the first-half or the second-half responses. The average agreement was .77, .78, and .78 respectively when the positive and negative item subscales were differentially weighted for second-half responses, for first-half responses, and for both first-half and second-half responses. Differentially weighting the positive and negative item subscales produced surprisingly little improvement, and this lack of improvement was consistent across all 13 areas of self-concept measured by the SDQ III. These results provide no support for the separation of any of the 13 SDQ III scales into positive and negative item subscales.

In Study 3 the verbal and mathematical achievement scores provide particularly relevant criteria for Math and Verbal self-concept responses. Marsh and O'Neill (1984; also see Marsh, 1986b) found that the corresponding measures of academic achievement and academic self-concept were substantially correlated, and that the agreement was quite specific to matching areas of academic achievement and self-concept. For present purposes, total scores were computed for all the Math and all the Verbal self-concept items, and subscale scores were computed for the corresponding positively and negatively worded items. The total scores were as highly correlated with achievement scores (.58 and .37 for Math and Verbal) as the

positive item subscales (.57 and .32) or the negative item subscales (.53 and .37). Furthermore, this agreement based on the total scores was virtually unimproved by differentially weighting the positive and negative item subscales (.58 and .38 for Math and Verbal). These results provide no support for the separation on the Math and Verbal self-concept scales into positive and negative item subscales.

Summary and Implications

The purpose of the present investigation was to develop more fully a construct validity approach to test whether the separation of positive and negative item subscales was substantively justified for responses to the SDQ III. No support for the separation of any of the 13 SDQ III scales was found in the prediction of self-concepts inferred by significant others, in the prediction of mathematical and verbal achievement, in the short-term stability of self-concept, and in the consistency of self-concept responses across split-half sets of items.

Several features of this research enhance the generality of the findings to other self-concept research and, perhaps, to other areas of personality research. First, the results were very consistent across the 13 SDQ III scales. Thus, the results are unlikely to be idiosyncratic to a particular area of self-concept or the wording of items used to infer a particular scale. Second, the construct validity approach developed in this investigation and the choice of criteria make it unlikely that the separation of positive and negative item subscales for the SDQ III would be justified by consideration of other validity criteria. Support for the separation based on the agreement between self-responses and responses-by-others, or on the prediction of mathematical or verbal achievement, would have provided strong support for the validity of the separation, but the lack of support based on these criteria is not a strong basis for inferring that support will not be found with other criteria. Support for the separation based on short-term stability and internal consistency criteria would not provide strong evidence for the validity of the separation. However, the lack of support based on these reliability-like criteria make it unlikely that support will be found for other criteria that can reasonably be described as validity criteria.

The construct validity approach developed here was used to test the validity of separating positively and negatively worded items designed to test the same self-concept scale. However, the approach has much broader application. Test construction experts recommend the use of positively and

negatively worded items for rating scales in personality, attitude and opinion measurement, and so the approach has applicability to these areas of research. Furthermore, rating scales constructed with positively and negatively worded items are just a special case of the bipolar rating scales that are broadly applied in personality research (e.g., masculinity-femininity, introversion-extroversion, internal vs. external locus of control). Recent research has demonstrated that such bipolar scales may consist of empirically distinguishable components, but the more important question is whether these differences are substantively important or merely the substantively irrelevant effect of a response bias artifact. Thus, the construct validity approach developed in the present investigation is particularly relevant for addressing these important questions.

Despite the advantages of the construct validity approach, a word of caution must be noted. When, as in the present investigation, differentially weighting positive and negative item subscales produces little or no improvement to the prediction of suitable validity criteria, there is no substantive support for the separation of the subscales. However, the interpretation of support for the null hypothesis is always dubious, and the generality of the findings to other contexts and other rating instruments must be done cautiously. When differentially weighting positive and negative item subscales does improve the prediction of validity criteria, support for the separation of the subscales must also be made cautiously as there may be viable alternative explanations. For example, Marsh's (1986b) study with responses by young children suggested that their responses to negatively worded items were influenced by reading skills. Thus, a negative item subscale of a Reading self-concept scale would probably correlate more highly with reading achievement than would a positive item subscale, but the support for the differential weighting may be due to a bias in the negative item subscale rather than its more accurate assessment of Reading self-concept. Similarly, if the validity criterion consisted of a self-report measure consisting of all negatively worded items, then a negative item subscale may be more highly correlated to it than would a positive item subscale. However, such support for the differential weighting might reflect a response bias that affected both the criterion and the negative item subscale. In summary, as is the case with all construct validity studies, interpretations should be based on theory and the accumulated empirical results from diverse studies that use different measurement instruments, different validity criteria, and different experimental approaches.

Footnotes

[1] -- The total scores used to represent each of the SDQ III factors in the present investigation were the unweighted sums of the responses to items designed to measure each factor. In the original analyses based on each of the three studies, factor analytically derived factor scores were used instead. Hence, some of the results reported here may differ slightly from those reported in the earlier analyses.

[2] -- The total score for each self-concept scale represents the unweighted (or, more accurately, the equally weighted) sum of responses to the positive and negative item subscales. Multiple regression and canonical correlation were used to determine the optimally weighted combination of the subscales. However, because an equally weighted combination of the subscales is one possible result of these empirical approaches, agreement on the unweighted total scores represents the absolute lower bound for agreement based on these empirical approaches. Because the improvement produced by differential weighting was so clearly trivial in all the analyses in the present investigation, the complicated issues of assessing statistical significance, capitalization on chance, and cross-validation of the empirical weightings were not considered. However, if the improvement due to differential weighting was not trivial, then these considerations would require further attention.

REFERENCES

Bachman, J.G. & O'Malley, P.M. (in press). Self-concepts, self-esteem, and educational experiences revisited (again). Journal of Personality and Social Psychology,.

Benson, J. & Hocevar, D. (in press). The impact of item phrasing on the validity of attitude scales for elementary school children. Journal of Educational Measurement.

Byrne, B. (1984). The general/academic self-concept nomological network: A review of construct validation research. Review of Educational Research, 54, 427-456.

Carmines, E. G. & Zeller, R. A. (1979). Reliability and validity assessment. Beverly Hills, CA: Sage.

Cohen, J. & Cohen, P. (1975). Applied multiple regression/correlation analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum.

Gaudry, E., Vagg, P., Spielberger, C.D. (1975). Validation of the state-trait distinction in anxiety research. Multivariate Behavioral Research, 20, 331-341.

Jackson, D. N. (1967). Acquiescence response styles: Problems in identification and control. In I. A. Berg (Ed.), Response set in personality assessment (p. 71-111). Chicago: Aldine.

Jackson, D. N. & Messick, S. (1958). Content and style in personality assessment. Psychological Bulletin, 55, 243-252.

Jackson, D. N. & Messick, S. (1961). Acquiescence and desirability as response determinants on the MMPI. Educational and Psychological Measurement, 21, 771-790.

Marsh, H. W. (1986a). The bias of negatively worded items in rating scales for young children: A cognitive-developmental phenomena. Developmental Psychology, 22, xxx-xxx.

Marsh, H. W. (1986b). Verbal and Math self-concepts: An internal/external frame of reference model. American Educational Research Journal, 23, xxx-xxx.

Marsh, H. W., Barnes, J. & Hocevar, D. (1985). Self-other agreement on multidimensional self-concept ratings: Factor analysis & multitrait-multimethod analysis. Journal of Personality and Social Psychology, 49, 1360-1377.

Marsh, H. W. & Hocevar, D. (1985). The Application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups.

Psychological Bulletin, 97, 562-582.

Marsh, H. W., & Jackson, S. (in press). Multidimensional self-concepts, masculinity and femininity as a function of women's involvement in athletics. Sex Roles, .

Marsh, H. W. & Myers, M.R. (in press). Masculinity, femininity and androgyny: A methodological and theoretical critique.Sex Roles, .

Marsh, H. W., & O'Niell, R. (1984). Self Description Questionnaire III (SDQ III): The construct validity of multidimensional self-concept ratings by late-adolescents. Journal of Educational Measurement, 21, 153-174.

Marsh, H. W., Richards, G., & Barnes, J. (1986). Multidimensional self-concepts: The effect of participation in an Outward Bound program. Journal of Personality and Social Psychology, xxx-xxx.

Marsh, H. W., Richards, G., & Barnes, J. (in press). Multidimensional self-concepts: A long term followup of the effect of participation in an Outward Bound program. Personality and Social Psychology Bulletin, .

Marsh, H. W. and Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. Educational Psychologist, 20, 107-125.

Naylor, F. D. (1978). Success and failure experiences and the factor structure of the state-trait anxiety inventory. Australian Journal of Psychology, 30, 217-222.

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton: Princeton University press.

Wiggins, J. S. (1973). Personality and prediction: Principles of personality assessment. Menlo Park, CA: Addison-Wesley.

Table 1

Correlations Between Various Combinations of Indicators of Each of the 13 Self Concept Facets Based on Positively-Worded Items (P1 & P2), Negatively-Worded Items (N1 & N2) and their Totals (T1 & T2) for Studies 1, 2 and 3.

	Correlations Relating:																	
	P1 with P2[a]			N1 with N2[a]			T1 with T2[a]			T1 with P2 & N2[b]			T2 with P1 & N1[b]			P1 & N1 with P2 & N2[c]		
Area of Self-concept	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Mathematics	78	91	84	68	86	74	75	91	85	76	91	86	77	91	86	78	92	86
Verbal	59	84	58	56	83	61	62	88	83	64	88	83	62	88	83	64	88	83
Academic	32	86	81	30	85	67	36	87	71	36	87	71	36	88	74	36	88	74
Problem Solve	42	83	63	41	80	31	46	86	65	47	86	66	48	86	67	51	87	67
Phys Ability	75	86	85	72	82	84	77	88	91	77	88	91	77	88	91	77	88	91
Appearance	50	83	75	36	81	63	48	85	75	49	85	75	48	85	79	51	85	79
Same Sex	46	83	55	39	78	35	45	85	61	45	85	62	47	85	61	47	85	62
Opposite Sex	51	85	82	46	86	80	51	89	86	52	89	87	52	89	88	52	89	88
Parents	72	85	69	71	84	65	76	87	70	76	88	71	76	88	70	76	88	71
Spiritual	78	94	85	78	91	54	82	93	85	82	94	86	82	94	85	82	95	88
Honesty	29	68	42	38	73	35	37	75	62	37	76	62	38	76	62	39	76	62
Emotional	54	75	63	63	85	70	63	85	81	63	85	82	63	85	81	63	86	83
General	41	85	74	41	86	75	44	88	82	44	88	82	44	88	83	44	88	83
Mean	54	84	70	52	83	62	57	87	77	57	87	77	58	87	78	58	87	78
SD	17	07	14	17	05	16	16	04	10	16	04	10	16	04	10	16	04	10
Median	51	83	74	46	83	65	51	87	81	51	88	82	52	88	81	52	88	83

Note. All correlations, presented without decimal points, are statistically significant ($p < .001$)

[a] coefficients in these columns are simple bivariate correlations.

[b] coefficients in these columns are multiple regressions relating a total score to the optimally weighted combination of a positive and negative item subscales.

[c] coefficients in this column are canonical correlations relating a the optimally weighted combination of one set of positive and negative item subscales to the optimally weighted combination of the second set of positive and negative item subscales.